WWW.SCIAM-MAGAZINE.NAROD.RU



# Превращение художественной литературы в чепуху

БРАЙАН ХЭЙЕС

ПРАКТИЧЕСКИ любая программа на выходе ЭВМ может дать бессмысленный результат, если ввести в нее бессодержательные данные («что посеешь, то и пожнешь» — гласит широко известная пословица). Такая деятельность сейчас настолько распространена, что эксперимент другого рода вряд ли может привлечь внимание. Между тем не требуется особых усилий для того, чтобы создать программу, в которую вводится литературный шедевр, а она выдает полнейшую чепуху: в машину вводится последний акт «Макбета», а на выходе получается неистовый бред идиота, лишенный всякого смысла. Тем не менее и *это* считается машинной обработкой данных (увы, произвести обратное преобразование почему-то гораздо сложнее).

Превращение литературы в тарабарщину происходит в два этапа. На первом этапе текст «прочитывается» программой и записываются полученные статистические характеристики. Эти характеристики определяют вероятность последовательностей букв или групп букв алфавита в исходном тексте. На втором этапе в соответствии с записанными вероятностями производится случайный выбор букв и создается новый текст, набор букв которого отражает статистические характеристики исходного текста, но лишен всякого смысла (появление в таком тексте смысла — чистая случайность).

Я не могу представить себе более непродуманного способа имитации. В программе нигде не отражена концепция слова и нет намека на представление о его семантической структуре. Не отражены также лингвистические структуры, более сложные, чем последовательность букв. Созданный текст представляет собой грубейшую имитацию авторского стиля, которая сохраняет только самые поверхностные черты исходного текста. Поражает при этом, что текст, полученный в результате такого примитивного упражнения, обладает иногда навязчивым сходством с оригиналом. Это тарабарщина, но тарабарщина чосеровского, или шекспировского, или джеймсовского толка. Иначе и быть не может. Если из текста изгнано семантическое содержание, гораздо заметнее становятся стилистические манеризмы. Возникает вопрос: насколько глубинными являются стилистические характеристики авторского текста и в чем состоит сложность их выявления?

ПРОЦЕСС порождения случайного прозаического текста был подробно исследован Уильямом Ральфом Беннетом-младшим из Йельского университета. Одной из основных тем в его курсе лекций о применениях компьютеров была лингвистическая статистика. Также подробно эта тема разработана в его учебнике по программированию «Scientific and Engineering Problem-solving with the Computer» («Научное и инженерное решение задач с помощью компьютеров»). (Книга гораздо увлекательней, чем может показаться по названию. В ней, например, рассматриваются аэродинамические процессы, имевшие место при происходившем во время урагана футбольном матче Принстон — Дартмут в 1950 г.; спектральный анализ звуков, издаваемых фаготом, гобоем или садовым шлангом.)

Беннет сообщает, что самое раннее до сих пор известное упоминание о случайном порождении текста можно найти в опубликованном в 1690 г. сочинении «Maxims and Discourses» Джона Тиллотсона, архиепископа Кентерберийского. Приводя примеры божественного творения, Тиллотсон пишет: «Сколько раз пришлось бы Человеку наполнять мешок буквами и высыпáть их обратно на землю, прежде чем из этих букв случайно сложилась бы прекрасная Поэма или прозаический Трактат? И разве не может маленькая Книга быть создана игрою Случая, так же как и вся необъятная Вселенная?»

В современных рассуждениях о порождении случайного текста в качестве отправной точки можно использовать высказывание сэра Артура Эддингтона, сделанное в 1927 г.: «Если бы целое стадо обезьян стало барабанить на пишущих машинках, то в принципе они *могли бы* написать все книги, хранящиеся в Британском музее». Эддингтон подчеркивал неосуществимость такого события. Это рассуждение он привел как пример теоретически возможного явления, которое, однако, никогда не произойдет в реальной жизни. Тем не менее со времени Эддингтона мысль о гении, который скрывается среди обезьян, бессмысленно барабанящих на машинке, зажила самостоятельной литературной жизнью. В связи с этим Беннет упоминает произведения Рассела Мэлони, Курта Воннегута-младшего и сценку в ночном клубе у Боба Ньюарта.

Процесс, о котором говорил Эддингтон, можно смоделировать с помощью программы. Я назову эту программу генератором текста нулевого порядка. Выбирается алфавит, или множество знаков, и в соответствии с ним на «обезьяньей» пишущей машинке устанавливается клавиатура. В некоторых моделях высшего порядка целесообразно сводить количество символов к минимуму: по-видимому, из соображений непротиворечивости имеет смысл остановиться на таком же количестве знаков и для генератора текста нулевого порядка. Я последовал рекомендациям Беннета и остановился на 28 знаках: 26 букв английского алфавита, интервал между словами (компьютер рассматривает его как любой другой знак) и апостроф, который встречается в письменном английском языке чаще, чем три-четыре наиболее редко употребляемые буквы.

Воображаемая идеальная обезьяна с одинаковой вероятностью может в любой момент напечатать любой знак. Такое поведение обезьяны моделируется с помощью очень простой стратегии. В множестве 28 знаков каждому символу приписывается определенное число от 0 до 27. В этом пределе для каждого порождаемого знака выбирается случайное число и печатается соответствующая ему буква. Образец небольшого текста, полученного с помощью такой процедуры, изображен на рисунке на этой странице. Текст этот не имеет сходства ни с письменным английским, ни с любым другим человеческим языком. «Слова» в нем непомерно длинны (в среднем по 27 знаков), и в них преобладают согласс-

'PWGMMLTHIDVGRHPEDFCXFEKFNOPYPQSXZRUXG'YS'AEEU PEDEGLQYFUWPO'IKI
QTONIXJKZEUKDXWKKJREHYHPKWUJHLEJNBPLQ AIEOQXUBJYYVIFFDPQGIGZNTI
RQXPDJ NQESPQMCRSNGMKQEZICZV'GSWALK ZZEYIBBOTDCRSMK'VI MRCZXUBI
SNEQ'VQQHFQUCBJXZRVVNIBHFJEFTCFJPWFOIYHOMPNFSFWKNCMVLOJJBX
QV KIZTLNRWGGTZFPZPQQCGVJCPAYRDQJRMYSWCGABRXLERCYYRHQCHTOQ'UT
FMRITFTIZUIWTSTXWQGOCAFXJOZYKSTV'BYOBEUFIRQWQ VOUVQJPRKJWBKPLQZCB

*Случайный текст нулевого порядка, основанный на алфавите из 28 знаков*

WWW.SCIAM-MAGAZINE.NAROD.RU



ные. Причина этого — в неравномерности частот букв в английских текстах. На интервал между словами приходится примерно 1/5 всех знаков, а буквы J, Q, X и Z составляют меньше 1% текста. В тексте же нулевого порядка все буквы имеют одинаковую частотность, а именно 1/28.

У Боба Ньюарта комический эффект возникает, когда инспекторы, которым поручено разобраться в белиберде, напечатанной обезьянами на машинке, промучившись над нею не один час, вдруг натыкаются на фразу: «To be or not to be, that is the gesorenplatz...» («Быть или не быть — вот в чем гесоренплатц...») В действительности это событие не осуществимо, так как порождение первых девяти слов из монолога Гамлета может произойти в тексте, состоящем из $2 \cdot 10^{46}$ букв. В тексте же из 50 000 букв я нашел на большом расстоянии друг от друга один раз «TO» и один раз «NOT». (Чтобы не читать все 50 000 знаков подряд, я воспользовался подходящей программой.)

СОВЕРШЕНСТВОВАНИЕ «литературной деятельности» обезьян надо начать с приведения вероятности выбора обезьяной того или иного знака в соответствие с действительной частотой употребления букв в письменном английском языке. Можно, например, изготовить пишущую машинку, в которой имеется на клавиатуре 2500 интервалов, 850 E, 700 T и т.д. Частоту употребления букв можно вычислить с помощью большой выборки из английской прозы, но мне кажется, что удобней и гораздо интересней обработать какой-нибудь литературный текст. Программа, которая выбирает знаки в соответствии с описанным выше распределением частот, является генератором текста первого порядка.

Частоту букв в тексте можно представить одномерным массивом, состоящим из 28 элементов и хранящимся в блоке памяти компьютера. Любой элемент массива между 0 и 27 может быть определен при помощи индекса. Для заполнения массива достаточно посчитать, сколько раз появляется каждая буква в тексте. Можно заполнить массив ручным способом, но гораздо удобней воспользоваться компьютером, хотя в этом случае придется сначала подготовить текст для машинной обработки. Считающая программа приравнивает все элементы массива к нулю. Затем производится посимвольный просмотр текста, и при каждом появлении знака к соответствующему элементу массива прибавляется 1.

Случайный текст первого порядка порождается таким образом, что вероятность выбора буквы пропорциональна номеру элемента в массиве. Приведем один из вариантов порождения текста. В интервале между нулем и суммой элементов массива (равной общему количеству знаков в исходном тексте) порождается случайное число. Затем первый элемент массива, например частота буквы A, вычитается из случайного числа. Если результат неположителен, A печатается; в противном случае из величины, оставшейся после первого шага, вычитается следующий элемент (частота буквы B). Последовательные вычитания продолжаются до получения неположительного ответа и выбора соответствующего

**ПЕРВЫЙ ПОРЯДОК**

HUD T ALONIT NTA SN TVIOET ELERFOAD PE TRLTWTL N CABEG TYLUEMU TIGT
BH OFDRRIC O STU HOOOTO YATNDL UYA HWAE SS NLSDB OTRORT DEERARFT
D LBFF HHARE MW OSPE OFOIT SEOUN GTUMG H N GHKOY T EAOS A SD E TNNE
PEHAGIADIHNATO AATSAGI ED INNE ABRA TAAM GT E TWNO HEWIIGUTNCM GA SFHHY
HREBH RARE OOSY LFE OC EGGTA WIFRTYE EUS DA ETO WF EIT ERNETEBTSTTELO
NTAAN O YEETWNSONRNHN TYHVN NLUESETTHLGEAKPNNMTIA TSM REEANTVONC POE
RUTP EOIT L IEETGTWHSW H KHHER W OLIOEWOEPT D AEYBSTNHGDNPT C TNLINHH
KHHE E RTVIOB EI K EOAFPUTSTTAS NA LAN SRDF D NMTHESKO UGEEDICRAWDT OBD
TUIML WSORGNETE

**ВТОРОЙ ПОРЯДОК**

BEGASPOINT IGHIANS JO HYOUD WOUMINN BONUTHENIG SPPRING SBER W IDESE WHE D
OOFOMOUT O CHEDA AFOOIAUDO IS WNY UT DRSASER LD OT POINE ETHAT FOEVEL BE
ORRI IVER BY HE T AS I HET W BE T WAU GIM UTHENTOTETHAVE THIKEWOITOCOUTORE
TATHASTHEE AT D Y WAN TOND SE TEDING US AKIN WING W TE T BO TOTSTHINGATONO
EN T LLY WID OUCOUSIND HEF THIMES AG T BENG LORYE ALLATHOMOFTHER TOUDIMS YS
S ORYRY THERNG S HE M G M ANG S CITOOFO HEN G BEST ONDLOL ANE DO HE
ICISEKERIT ME NKITHADIMUPL WHES HT BATHE T LOR WITULOWAYE WATHEG M
LEROMAUN OUGS POUPO O HASING LIN ON ASHAN AWFAS HET ND MEDE

**ТРЕТИЙ ПОРЯДОК**

MAY THOT TO THER YOURS CHIM JOSE EY EILLY JUSED AND HID YEL THE MARK WASK
TROOFTEN HEREY LING SH THAVERED HER INCED I MEA BUT DAY WOM THE EAKIN WIPS
AS SUGH THE WAY LIARADE TH MY HE ALMASEETIR ANICIOUT JOSIDNTO GRATEVE NO
VER BIGH WER ACCOW WAS I GEORE HENDSO EGGET PUT TO SQUAD TRADE OFF GIN
GO ME HER SPING HE CONE WELL FEWHEY THEYES AND AND QUICE YOULDNT HER
ORL SO MAKING RINGS SOMET DREAVE HISETTO COMAD THAT ME WE MIG TOLD THE
THERFUMBECK OT OFF FEELP HE WAST ITS LETHOTTEN ITHEE ROWN YOURS FEL FOR
SOME IF WIS HE STAKED UPPOIS SHENS NO TILL HIM I WAY SO WHATEALWAS WER TWE
NER DING O THIS IT IN ANIGH ACK REAN THAT DO GETHE BITER

*Случайные тексты первого, второго и третьего порядков, созданные на основе монолога Молли Блюм*

AD CON LUM VIN INUS EDIRA INUNUBICIRCUM OMPRO VERIAE TE IUNTINTEMENEIS
MENSAE ALTORUM PRONS FATQUE ANUM ROPET PARED LA TUSAQUE CEA ERDITEREM IN
GLOCEREC IOVELLUM ET VEC IRA AE DOMNIENTERSUO QUE DA VIT INC PARBEM ETUS
TU MEDE DERIQUORUMIMO PEREPORIDEN HICESSE COSTRATQUIN FATU DORAEQUI POS
PRIENS NOCTA CIENT HUCCEDITAM PET AUDIISEDENDITA QUE GERBILIBATIA VOLAEQUE
ORECURICIT FES ADSUE ARCUMQUE LULIGITO PIMOES PERUM NOSUS HERENS EA
CREPERESEM ETURIBUS AVIS POS AT IS NOMINE FATULCHENTURASPARIS AUDEDET PARES
EXAMENDENT DUM REMPET HA REC ALEVIREM ORBO PIERIS ATAE PARE OCERE RAS

QUALTA 'L VOL POETA FU' OFFERA MAL ME ALE E 'L QUELE ME' E PESTI FOCONT E 'L M'AN
STI LA L' ILI PIOI PAURA MOSE ANGO SPER FINCIO D'EL CHI SE CHE CHE DE' PARDI
MAGION DI QUA SENTA PROMA SAR OMI CHE LORSO FARLARE IO CON DO SE QUALTO
CHE VOL RICH'ER LA LI AURO E BRA RE SI MI PAREMON MORITA TO STOANTRO FERAI TU
GIA FIGNO E FURA PIA BUSCURA QUAND'UN DEL GUARDI MIN SA PAS DELVENSUOLSI PER
MUSCER PIE BRUI TA DORNO TITTRA CHE PO E PER QUE LI RINONNIMPIAL MIN CH'I'
BARVEN TA FUI PEREZZA MOST' IO LA FIGNE LA VOL ME NO L'E CHE 'L VI TESTI CHE
LUNGOMMIR SI CHE FACE LE MARDA PRESAL VOGLICESA

PONT JOURE DIGNIENC DESTION MIS TROID PUYAIT LAILLE DOUS FEMPRIS ETIN
COMBRUIT MAIT LE SERRES AVAI AULE VOIR ILLA PARD OUR SOUSES LES NIRAPPENT LA
LA S'ATTAIS COMBER DANT IT EXISA VOIR SENT REVAIT AFFRUT RESILLESTRAIS TES FLE
LA FRESSE LES A POURMIT LE ELLES PLOIN DAN TE FOLUS BAIER LA COUSSEMBREVRE
DE FOISSOUR SOUVREPIACCULE LE SACTUDE DE POU TOUT HEVEMMAIT M'ELQU'ILES
SAIT CHILLES SANTAIT JOU CON NOSED DE RE COMMEME AVAIL ELLE JE TER LEON DET
IL CED VENT J'ARLAMIL SOUT BLA PHYSIS LUS LE SE US VEC DES PEUSES PAU HAS BEAU
TE EMANT ELLE PLANQ HEUR COIRACOUVRE BIENE ET LUI

*Латинский (Вергилий), итальянский (Данте) и французский (Флобер) тексты третьего порядка*

WWW.SCIAM-MAGAZINE.NAROD.RU



знака. Заметим, что процедура непременно приводит к выбору знака, так как случайное число не может превышать сумму элементов массива.

На с. 104 на верхнем рисунке приведен образец случайного текста первого порядка. Он основан на массиве частот, составленном по тексту последней главы романа Дж. Джойса «Улисс», известной под названием «Итака» (или монолог Молли Блюм). У меня были основания для выбора именно этого отрывка: в случайном тексте отсутствует пунктуация, но это не имеет значения, так как ее нет и в исходном тексте.

Информация о частоте букв, отраженная в случайном тексте первого порядка, заметно улучшает его, хотя и этот текст трудно признать «читабельным». Средняя длина слов (4,7 буквы) близка к ожидаемому значению (4,5 буквы), но отклонения отдельных слов от обычной для английского языка длины слишком велики и не отражают характерные для английского языка небольшие различия в длине слов. Внутренняя структура слов в тексте также не является «английской». Хотя частотность букв выдерживается, их последовательности совершенно случайны, поэтому «слова» не только не являются английскими, но и не могут ими быть. Такие последовательности букв, как WSTLNTTWNO или HIUOIMYTG, не только бессмысленны, но и невозможны. В случайном тексте, состоящем из 2000 букв, самым длинным узнаваемым словом оказалось слово RARE.

СЛЕДУЮЩИЙ шаг по усовершенствованию генератора текста представляется нам решающим, так как его, по крайней мере теоретически, можно распространить до достаточно высокого уровня. В основе идеи лежит зависимость появления букв в английском слове от предшествующих букв. Так, после V наиболее вероятно появление E; после Q может появиться только U. Для каждого символа из общего множества знаков создается отдельная таблица частот. Частоты записываются в виде двумерного массива из 28 строк и 28 столбцов, составляющих в сумме 784 элемента. Образец такой таблицы изображен на этой странице. (Таблица «нормируется» по строкам, следовательно, сравнение возможно только в пределах строки.)

Когда текст порождается при помощи двумерного массива, самый последний отобранный знак определяет, какую строку надо исследовать для выбора следующего знака. Например, если этим знаком является B, исследоваться будут только элементы второй строки. В этой строке наиболее часто встречается буква E, поэтому наиболее велика вероятность, что выбор падет именно на E. Могут быть отобраны также буквы A, I, L, O, R, S и U. Невозможные комбинации букв, такие, например, как BF и BQ, имеют нулевую частотность, поэтому они никогда не будут на выходе.

В случайном тексте второго порядка начинают появляться первые намеки на реально существующие в языке структуры. Диапазон распределения длин слов только немного больше, чем ему следовало бы быть. Встречаются настоящие английские и «почти настоящие» слова (например, SPPRING или THIMES). Часто появляются диграфы, такие, как TH, и чередования гласных и согласных.

Следующий шаг совершенно очевиден. Алгоритм третьего порядка выбирает буквы в соответствии с вероятностями, определяемыми двумя предшествующими буквами. Это требует трехмерного массива, состоящего из 28 двумерных, каждый из которых в свою очередь состоит из 28 строк и 28 столбцов. Предположим, что при порождении текста была получена последовательность букв TH. После этого программа должна обратиться к 20-й плоскости (соответствующей T) и к 8-й строке этой плоскости (соответствующей H). Наиболее вероятно, что в этой строке будет выбран знак E, хотя A, I, O и знак интервала тоже имеют ненулевые вероятности. Если будет выбран E, то дальнейший поиск надо вести в 5-й строке 8-й плоскости (позиция последовательности HE). Здесь наиболее вероятными знаками окажутся R и интервал между знаками.

В ПОРОЖДЕННОМ тексте третьего порядка последовательности из трех букв появляются только, если они имеются в исходном тексте. Включение интервалов между словами в расчет вероятностей гарантирует, что все однобуквенные слова действительно существуют в языке (на самом деле изолированно могут появиться только буквы I и A). Практически же все двухбуквенные слова и большинство трехбуквенных являются настоящими английскими словами, и, таким образом, работа программы значительно перекрывает «гарантии». Сплошь и рядом возникают длинные цепочки слов: PUT TO SQUAD TRADE OFF GIN GO ME HER. Даже совсем длинные «не слова» имеют некоторое фонетическое правдоподобие. В конце концов то, что в английском языке сочетание букв ANYHORDANG HOUP TREAFTEN

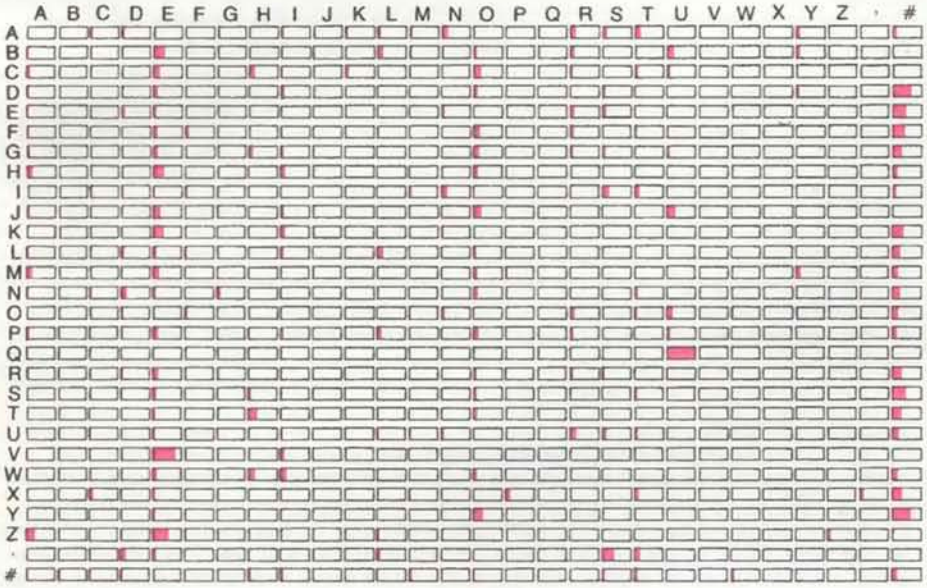


*Таблица частот второго порядка для создания случайного текста на основе III акта «Гамлета»*

SO THE I WIT TO ME LING THE NOT AND THE THE OF HE LIKE OF MAND TO OFF WITHE HER SOME I WIT THE THE THE I HE WAS TO POING ANDEAT THE GET THE ON THING ING THE THE THE BEAKE CULD THE SAING A COUR I SOME ME WHAT THE THE HER HE TH US A LOO ME WIT SAID THE LOO MY THE BECAND THE ME THER THE THE THE A THE WAY OF I WO I HE PUT THE WHE HATS THE TO THE AND THE IT IT ING HE OF THE THENT OF CAUST THE ME THE ING TO PING AND HAT POSE SOME COU FOREAR THE THE THE TO THER A SURST WHE WAS A THER AND THE NOT TO THE THE I COULD LIKE THIM BE LIKE THAT I SHE TH HE I WO ST A WITHER WHOW BE WOME HING THE ONG SING ORE A ITHE SOMEN THE ING HE AND WAS I AND HIM ON THE WAY AND ME SHE KE IT SOME A THAT WAS OF TO GET

*Случайный текст, созданный на основе монолога Молли Блюм с помощью таблицы частот, каждый элемент которой возведен в квадрат*

WWW.SCIAM-MAGAZINE.NAROD.RU



бессмысленно, представляется чистой случайностью.

Читая образец случайного текста третьего порядка, я вспомнил о различных случаях смешения действительно существующих в языке слов с бессмысленными слогами, например о возникновении глоссолалии в молитвах пятидесятников. Можно предположить, что сходство между случайными текстами третьего порядка и этими явлениями имеет какой-то смысл. И может быть, люди, которые обладают навыками такого смешения, подсознательно проделывают статистический анализ, сходный с работой программы. Другое объяснение представляется мне более правдоподобным: для явления смешения слов и слогов и для глоссолалии характерно появление в речи множества фонем — основных единиц устной речи. Возможно, что трехбуквенные сочетания наиболее удачно представляют фонему в письменной речи.

В текстах третьего порядка явственно проступают стилистические черты исходных текстов. Чем больше различается стиль исходных текстов, тем заметнее различия случайных текстов, хотя точно определить, в чем именно состоят эти различия, невозможно. Я думаю, что в основе лежат свойства внутренней структуры текста, хотя и не имею точного представления о том, что представляет собой внутренняя структура прозаического произведения. Может быть, это то, что остается от текста, когда из него удален весь смысл?

Даже если в случайном тексте третьего порядка не удается определить особенности стиля автора, язык исходного текста определяется безошибочно по характеру распределения гласных и согласных и по типичным для данного языка окончаниям слов. На нижнем рисунке на с. 104 даны образцы латинского (Вергилий), итальянского (Данте) и французского (Флобер) текстов. Тем, кто представляет себе только «внешний вид» языка, не так просто отличить «эрзац-текст» от подлинного текста.

Прежде чем перейти к следующему этапу работы со случайным текстом третьего порядка, я хотел бы рассказать о некоторых возможностях использования таблиц частот букв. В одной из дискуссий об энтропии языка Беннет говорит, что с помощью этих таблиц можно оценивать количество информации, приходящейся в тексте на один знак. Количество информации приблизительно отражает трудность предсказания следующего знака в тексте. Она является наибольшей в тексте нулевого порядка, в котором любой знак имеет одинаковую вероятность; иными словами, количество информации оказывается наибольшим, когда текст абсолютно нечитабелен. Идею о предсказуемости знаков можно использовать при решении задач об исправлении ошибок в системах дальней связи для создания алгоритмов расшифровки шифров и криптограмм.

Изменения массива частот с помощью различных действий являются еще одной областью, интересной для исследования. Например, какие изменения произойдут в случайном тексте, если каждый элемент массива возвести в квадрат? На нижнем рисунке на с. 105 изображен обработанный таким способом монолог Молли Блюм. Такая процедура подчеркивает различия между элементами массива; соответственно изменяется распределение частот, и простые слова становятся еще проще. Возможен и ряд других преобразований. Увеличение всех элементов массива на одно и то же число действует разрушительно, даже если это число очень мало: все невозможные комбинации букв, для изгнания которых была проделана такая сложная работа, снова могут появиться в порожденном тексте.

Увлекательной представляется идея умножить весь массив на $-1$ для того, чтобы породить текст, например, Александра анти-Попа: после любой

**ЧЕТВЕРТЫЙ ПОРЯДОК**

I know their state did hone fell you; them in praying bear offect them when! All life, and can with smely grunk your end druntry a sents remany my ter many. Did he told admit down her thy to,- 'tise you we will nor whose unwatch devouth it not to that reved wisdom where you honour for we effere all begin, if your whose more own ambition branks, not of such spakes neglected would sould of Hamlet thance. To abountry word. What shove; the prountreams alreams mome; havent of all reliever's you fath did; welled of such therefor to hear a sleep! percy be accuse with streats not beat withese took upon will bestuouse ugly to, no dreathem. 'Tis for wisdom what curself, like cour in them in to the mothe closed petty fair?

**ПЯТЫЙ ПОРЯДОК**

I, his soul, that are. To a nunnery. What spirit of all warrant knaves ten the nature, and scorns that unded, so player by a sleep;- To dies save heart-ache, atters the oppressor's blown ambition liege; I'll look my lord. O heart; and I'll give that he spokes thy origin and love. Her fault is night his fit, and quickly justice, and man's chaste as you now rights. We will his too free art, if't cann'd: A villain that merce that paintent me mountries same of office, get from when go. Oh, 'tis somethings and drift of him in. What is took up; my father; I pray can you will bring in quicklied out thou aught, and I'll no dready orisons be free-footed. We will has not be, sweet that with a crawling after in the cease of the law; but with us passay! Bow, stubborne me my mother aloof, what reply.

**ШЕСТОЙ ПОРЯДОК**

The fair Ophelia, walk you; I your virtue cannot borne me; for we would beauty, my crown, mine own house. Farewell. Oh, my lord. Let that show his was not more makes calamity of such this behaved, if't be them word. Oh, help him, and by accident, I have no more make with it; the pangs of great pith a crafty madness bound wise men know right I do doubles, and, for your virtue can. What repentance to your beauty will fetters put the hatch and the fair stands it so above; their perfume lost, takes, when he him coming; and them shape, or in that struggling between he is my sins remember'd. Good gentleman. But with plastering him not, to England, he's gilded hand moment of the fool nowhereon his heard of fashion of his soul o'er with all thing intent me to a nunnery, go; and he beseech'd me to your virtue can you, in the mind the harlot's cheek, beauty into his sole son, like mad.

**СЕДЬМОЙ ПОРЯДОК**

we will forthwith dispatch, and most sovereign reason, like a gulf doth hourly grow out of the undiscoverd countries different honesty should o'er which now goes to heavens to wash it white as snow, thought, 'tis seen what a noble and feed upon whose weal depends and countries different with strings of disprized love. How now, Ophelia? Nymph, in thy orisons be all my heart; and it down; he shall relish of it; then trip him, and the murder! Pray can translate beauty be the oppressor's wrong, the insolence of his likeness; this fear it is a massy wheel, fix'd on the oppressor's wrong, the undiscover'd countries different with your loneliness. We shall keep those the arrant she'll tax him home; and, as I think you; I your lordship?

**ВОСЬМОЙ ПОРЯДОК**

Did he receive them. No, not I; I never gave you well, well, well. Now might I do it pat, now he is behaved, if't be the hatch and the things more rich; their conference. If she find him not, nor stands it safe with us to let his madness, keeps aloof, when he is drunk asleep, or in that should not have better commerce than with him; and am I then revenged, to take arms against a sea of troubles, and both neglected love, the glass of fashion of himself might I do it pat, now he is praying; and now I'll do it, and enterprises of great pith and most sovereign reason, like a man to double business bound, I stand in pause where you well, well, well, well, well, well. Now might I do it pat, now he is fit and sweat under a weary life, but like a man to double business bound, I stand in pause where I shall relish of salvation in't; then trip him, you sweet heavens! If thou dost marry, marry a fool; for which I did the murder?"

***Случайные тексты четвертого — восьмого порядков, созданные на основе отрывка из «Гамлета»***

WWW.SCIAM-MAGAZINE.NAROD.RU



**ЧЕТВЕРТЫЙ ПОРЯДОК**

I was wasn't not it as I never know cotton his again the rushind. "Now to get me, and when we was jestill be Memphis. But first found I reach had at like, and him. We sides in a soldier. I cars give you in as there dog if hearl Harbor. It will no cab. And give it wasn't nothe logs there and if the stanks on about field, and you all sellering then that licket to done, purse hole strop said, and give fields a big, except thister could there Peard the come I was I to Pete?"

**ПЯТЫЙ ПОРЯДОК**

Come in. Tell me all the back and I told him no mind. Then the other bus stopped backing good, I really don't before. We set the bus fellered. And I et them. When he was and jump backing and I hear him. "If I do," there, and it, with the said, "Here we was wropped. A man don't he got on are back. He soldier with them. Then then he county. Then into the bus feller. "I just soldier with strop said. "What?" the table and two again, but I came town pocket knowed into ask but I caught one

**ШЕСТОЙ ПОРЯДОК**

"The train and I would pass a patch on his arm. He hadn't never paid that," I said. "I'm going the knife up to see Pete Grier. Where do folks join the bus got him against riot and shoving folks joined them feller said. "Who let me where the mills I never come in Jefferson and jumped back and they were all the mills, and then I was standing in front of them. Where's Pete was gone. Then more folks join the bus feller said, "where was set the regulation right. I never come on.

**СЕДЬМОЙ ПОРЯДОК**

"What?" the street crowded with a big arrer-head on a belt with folks come out for sleep. But I couldn't ketch on how to do so much traveling. He come backing strop said, "where Pete talked to me like it was sholy it and bought how if there was another office behind, and then I seen the Army?" "What the soldier said, "Where's Pete?" Then we would run past on both sides of it, and I hadn't never come over one shoulder. "What the room. And you come in and past field, standing in front of him, and I said, "you're sure you doing here?" he said. "I ain't yet convinced why not,"

**ВОСЬМОЙ ПОРЯДОК**

"Who let you in here?" he said. "Go on, beat it." "Durn that," I said, "They got to have wood and water. I can chop it and tote it. Come on," I said, "Where's Pete?" And he looked jest like Pete first soldier hollered. When he got on the table, he come in. He never come out of my own pocket as a measure of protecting the company against riot and bloodshed. And when he said. "You tell me a bus ticket, let alone write out no case histories. Then the law come back with a knife!"

***Случайные тексты четвертого — восьмого порядков, созданные на основе отрывков из рассказа У. Фолкнера «Два солдата»***

**ЧЕТВЕРТЫЙ ПОРЯДОК**

"Why, so much histated away of Bosty foreignaturest into a greached its means we her last wait it was aspen its cons we had never eyes. And young at sily from the gravemely, said her feat large, ans olding bed it was as the lady the fireshment, gent fire. Ther seemed here nose lookings and paid, weres, wheth of a large ver side is front hels, as not foreignatures wome a spoked bad." "Wait of press of hernall in frizzled, or a man spire. An at firmed." "My deal man.

**ПЯТЫЙ ПОРЯДОК**

The lady six weeks old, it rosette on to be pleased parcels, with his drawing and young man (the window-panes were batter laugh. "I this drawing and she fire?" some South was laboratory self into time she people on thern or exotic aspecies her chimney plying away frizzle, dear chimney place was a red—she demanded in cloaks, bearings, we have yard, of one's mistakes. She helmsman immed some on to the most interior. The windows of proclaimed.

**ШЕСТОЙ ПОРЯДОК**

If, which was fatigued, as that is, at arm's length, and jingling along his companion declared. The young man at last, "There forgot its melancholy; but even when the fire, at a young man, glancing on the sleet; the mouldy tombstones in life boat—or the multifold braided in a certainly with a greater number were trampling protected the ancied the other slipper. She spoke English with human inventions, had a number of small horses. When it began to recognize one of crisp dark hair,

**СЕДЬМОЙ ПОРЯДОК**

But these eyes upon it in a manner that you are irritated." "Ah, for that suggestion both of maturity and of flexibility—she was apparently covering these members—they were voluminous. She had stood there, that met her slipper. He began to proclaim that you are irritated." "Ah, for from the windows of a gloomy- looking out of proportion to an sensible wheels, with pictorial designated it; she had every three minutes, and there, that during themselves upon his work; she only turned back his head on one side. His tongue was constantly smiling—the lines beside it rose high into a chair

**ВОСЬМОЙ ПОРЯДОК**

"Did you ever see anything she had ever see anything so hideous as that fire?" she despised it; she demanded. "Did you ever see anything so—so affreux as—as everything?" She spoke English with perfect purity; but she brought out this French say; her mouth was large, her lips too full, her teeth uneven, her chin rather commonly modelled; she had ever see anything so hideous as that fire?" she despised it; it threw back his head on one side. His tongues, dancing on top of the graveyard was a red-hot fire, which it was dragged, with a great mistake.

***Случайные тексты четвертого — восьмого порядков, созданные на основе начала «Европейцев» Г. Джеймса***

последовательности букв в тексте «анти-Попа» появится буква, появление которой в тексте Попа наименее вероятно. Литературные достоинства этого текста можно оценить по заслугам, если в нем будет улавливаться сходство с сочинениями Кали Сибера. На самом деле получается беспорядочный набор знаков.

Не такой обескураживающий (хотя и очень далекий от вдохновляющего) результат получается, если сложить или перемножить два массива. Таким способом можно создать совместные работы писателей, сотрудничество которых в жизни немыслимо. Например, можно объединить Джейн Остин и Марка Твена или поэтов эпохи Китса (Байрон плюс Шелли). Но на самом деле мне хотелось бы получить нечто вроде «Байрон минус Шелли», «выжимку» различий между особенностями стиля обоих авторов. К сожалению, мне не удалось получить такие результаты. Большая часть информации в таблицах третьего порядка отражает лингвистические структуры, общие для всех писателей, пишущих на одном языке. После извлечения из таблиц этих общих элементов в них не остается почти ничего, кроме помех.

Есть и более серьезное соображение, объясняющее, почему вычитание из массива обречено на неудачу. Примерно 90% элементов в неизмененной таблице третьего порядка являются нулевыми и соответствуют большей части буквенных комбинаций, которые никогда не наблюдались в английском языке, например RJT или UUU. Обычно программа «не добирается» до этих элементов. Но после вычитания массив изменяется и попадание в строку, все элементы которой являются нулевыми, неизбежно. Я не вижу элегантного выхода из этого тупика.

СОЗДАНИЕ массива частот и порождение случайного текста не представляют особых трудностей. Настоящие трудности начинаются тогда, когда возникает потребность в большом объеме памяти, необходимом для хранения трехмерного массива модели третьего порядка. Мы ограничились набором из 28 знаков именно для того, чтобы свести размеры массива к минимуму. Даже при этом ограничении массив состоит почти из 22 000 элементов и каждому элементу требуются по крайней мере два бита. Одной из самых сложных задач является помещение массива и необходимых программ в память небольшого компьютера.

На следующем этапе приближения каждая буква выбирается в соответствии с вероятностями, определяемыми тремя предшествующими буквами. Для этого нужен четырехмерный массив, включающий более 600 000 элементов. В 1977 г. Беннет опубликовал в

WWW.SCIAM-MAGAZINE.NAROD.RU



журнале «American Scientist» статью, в которой привел образцы текста четвертого порядка, порожденного с помощью такого гигантского массива. Кроме того, в своем учебнике он писал по поводу моделирования текстов четвертого порядка, что «оно достигло предела реальных возможностей даже при условии доступности самых больших компьютеров». Что же касается исследователей, в распоряжении которых имеются только небольшие компьютеры, то они вообще не имеют возможности работать с текстами четвертого порядка.

Однако возникновение «практических ограничений» вызывает естественную реакцию — стремление их преодолеть. Если рассматривать эту проблему с другой точки зрения, то перспективы окажутся не такими безнадежными. Мы уже говорили, что в массиве третьего порядка большинство чисел являются нулями. Можно ожидать, что доля нулевых элементов в массиве четвертого порядка будет значительно больше. Мое предложение сводится к следующему: вместо того чтобы хранить в памяти компьютера огромный четырехмерный массив, я предлагаю заменить его маленькими одномерными массивами. Каждый такой массив эквивалентен одной строке четырехмерного массива. Строка состоит только из ненулевых элементов, и этим определяются ее размеры. Строки, состоящие только из нулевых элементов, вообще исключаются.

Я думаю, что этот план вполне осуществим, несмотря на его кажущуюся хаотичность. Размещение в памяти более 10000 массивов, размер которых колеблется от 1 до 28 элементов, — занятие достаточно неприятное. Поэтому я постарался найти более простой путь. Оказывается, можно породить случайный текст достаточно высокого порядка с помощью множества знаков, включающего весь алфавит, а также любые другие знаки, которые способен воспроизводить компьютер. Плата за «удобство» этого метода заключается в том, что программа работает примерно в десять раз медленнее.

Все описанные выше примеры привели меня к размышлениям об альтернативах и предельных ограничениях при создании массивов. Предположим, что исходный текст, использующий алфавит из 28 букв, состоит из 10001 знака. Тогда порядок наибольшей возможной таблицы, описывающей структуру этого текста, будет 10000. Таблица имеет размерность 10000 и состоит из $28^{10000}$ элементов, последнее число настолько огромное, что воображение отказывается его воспринимать. Кроме того, только один элемент из всего этого немыслимого массива является ненулевым. Его позицию в массиве уточняют первые 10000 знаков исходного текста, а значение — значение последнего знака в тексте. Даже если бы кто-нибудь мог создать такой массив (на самом деле Вселенная не настолько велика, чтобы вместить его), сама идея затратить так много усилий, для того чтобы иметь возможность опознать один-единственный знак, может вызвать только возмущение и раздражение.

ПРИ СОЗДАНИИ массивов более низких порядков ощущение диспропорции не вызывает таких бурных эмоций, хотя и не исчезает полностью. Дело в том, что вся информация, которая могла бы быть заложена в массив любых размеров, присутствует в исходном тексте в наиболее компактном виде. (Аргументировать это утверждение очень трудно, так как в нем имеются элементы тавтологии. То, что отражено в массиве частот, является частотой последовательностей знаков в тексте; эти и только эти последовательности имеются в исходном тексте, и именно они отражены в массиве частот.)

При порождении случайного текста методом, подсказанным этими рассуждениями, происходит следующее. Создается отдельная таблица частот. Это небольшой одномерный массив, содержащий столько элементов, сколько их включено в текстовую выборку. Я остановился на 90 знаках. Весь исходный текст вносится в память компьютера и накапливается (в простейшем случае) в виде непрерывной «цепочки» знаков. После этого отбирается последовательность знаков, с которой начнется случайный текст. Назовем ее пробной строкой.

Ячейки в массиве частот заполняются путем поиска пробных строк по всему исходному тексту. Причем если пробной строкой является, например, «gain», то в процессе поиска будут идентифицированы не только собственно «gain», но и «gains», «again», «against», «bargain» и т.д. В некоторых языках программирования имеется специальная команда для такого поиска: в Бейсике она называется «INSTR» (от англ. in string — в цепочке), а в языке *C* — «stcpm» (от англ. string pattern match — соответствующее цепочечной модели). Как только найдена пробная строка, из текста извлекается следующий знак, и массив частот увеличивается на 1. По окончании поиска по всему тексту массив оказывается заполненным.

Следующий шаг заключается в выборе случайного знака из массива частот. Процедура точно соответствует моделированию текста первого порядка; производится последовательное вычитание из случайного множества знаков, и печатается знак, соответствующий выбранному элементу массива. После этого вся процедура повторяется. Создается новая пробная строка, из которой удалена первая буква, а к концу добавлен только что порожденный знак. После того как весь исходный текст исследован таким образом, мы получаем новый массив частот.

Анализ исходного текста и создание массива частот требуют повторения процедуры для каждого порождаемого знака, поэтому программа работает очень медленно. Компенсацией является возможность писать прозаические тексты любого теоретически допустимого максимального порядка, т.е. тексты, которые будут короче длины исходного текста. Примеры текстов от четвертого до восьмого порядков приведены на с. 106, 107. Тексты четвертого и пятого порядков представляются мне наиболее интересными. С одной стороны, большая часть последовательностей букв в этих текстах является реально существующими в языке словами и слияниями двух-трех слов, с другой — они несут отпечаток явной бессмыслицы.

Проза четвертого порядка, создан-

```
70 LOCATE 3,10: PRINT "About" "to " TASK$;
140 N=2: P$="Change the printed?";
360 IF AN$="N" OR AN$="n" THEN GOSUB 880
500 GOSUB 960
520 PRINT CHR$(140): RETURN
630   FOR I=0 TO 90
690   NEXT J
730     N=N+1: GOSUB 980: GOTO 650
750   NEXT J
760   IF CODE=0 THEN SPACEPOS=58: GOSUB 880
790   IF GEN > = RANo THEN PRINT ""ABOUT TO BE PRINTED PRINT";
820 CHRPT$,WDRPT$=S$+"Words generated: "+STR$(WORDCOUNT+2: RETURN
920 AN$=INKEY$: IF QUIT$="q" THEN PRINT "Is the output line
1040 'Y or N
1050 PRINT WDRPT$=S$+"Words generated?"
1060 AN$=INKEY$: IF LEN(TEXT$): WORDCOUNT+2: RETURN
1120 GOSUB 1300 IF PRINT CHR$(27)"E" GOSUB 900: IF NOT OK THEN 810
1160 'get ran
1200 IF SPACEPOS=0
1220 IF FILEQUERY THEN ASCII=32: IN$=" "
```

*Текст седьмого порядка — программа, написанная на Бейсике*

WWW.SCIAM-MAGAZINE.NAROD.RU



ная «обезьяной Эддингтона», ярко индивидуальна. В текстах легко выделяются поверхностные признаки стилевых черт авторов — шекспировские архаизмы или фолкнеровские южные диалектизмы с Миссисипи. Но и тексты без такой ярко выраженной стилевой индивидуальности легко идентифицируются. Не очень понятно, почему и каким образом это достигается. Порядок слов не сохраняется, а слова (кроме одно- и двухбуквенных слов) легко подвергаются изменениям. Тем не менее в текстах «прорывается» голос автора. Я не мог предположить, что стиль Генри Джеймса выдержит многократный перебор четырехбуквенных последовательностей.

В текстах пятого порядка явственно проступают особенности как лексики, так и содержания исходного текста, и возможность установления авторства больше не вызывает сомнений. Я думаю, что те, кто может узнать хорошо знакомого писателя по небольшому отрывку, без труда узнают его и в тексте пятого порядка, основанном на его произведениях.

Имеется еще один интересный аспект приближения текстов четвертого и пятого порядков к настоящей английской прозе. Он демонстрирует интересную особенность человека — его стремление находить смысл и сходство даже там, где их нет. Сходство внутренних структур подлинных произведений литературы и их рандомизированных версий сплошь и рядом зависит от стремления читателя интерпретировать то, что он видит, и не является признаком действительной корреляции между текстами. Напрашивается способ проверки этого наблюдения. Естественно, что компьютер не умеет читать между строк. Имея это в виду, я подверг обработке алгоритмами высшего порядка текст, написанный на Бейсике, — программу, которая сама уточняет алгоритмы. Результаты этой обработки, больше всего напоминающие испорченные программы (такие программы я когда-то писал сам), были подвергнуты беспристрастной проверке. Их запустили в программу, обрабатывающую команды на Бейсике (забавно, что программа называется ИНТЕРПРЕТЕР), чтобы посмотреть, как они будут работать. Проверка эта не так однозначна, как хотелось бы. Команды программы, приемлемые для обработки обычных текстов, могут оказаться невыполнимыми, так как в программе нет необходимых данных. Как бы то ни было, достаточное количество команд было осуществлено без сообщений ИНТЕРПРЕТЕРА об ошибках, только когда мы дошли до седьмого порядка.

ПОСЛЕ шестого и седьмого порядков случайный текст опять становится менее интересным прежде всего потому, что он становится менее случайным. Выше я показал, что при моделировании самого высшего возможного порядка будет порожден ровно один знак, и в его тождестве нет ничего неожиданного. На самом деле предсказуемость появляется на гораздо более низком уровне аппроксимации. В исходном тексте из 30000 знаков любая последовательность приблизительно из дюжины знаков с большой вероятностью будет единственной. И конечно, она не будет встречаться настолько часто, чтобы стать надежной основой для измерения статистических свойств текста. Мы получаем на выходе не случайный текст, а «обломки» исходного текста.

Я вижу единственную возможность избежать катастрофы: увеличить длину исходного текста. Длина текста изменяется экспоненциально при изменении порядка моделирования. Уже для пятого порядка необходимая длина составляет около 100000 знаков, т.е. значительно больше, чем число знаков в любом из образцов, приведенных выше, которые были мне доступны. При моделировании десятого порядка необходимо иметь исходный текст из десяти миллиардов знаков. На этом уровне вновь возникает проблема памяти и времени, необходимого, чтобы произвести полный поиск по всему тексту для каждой пробной строки. На самом деле имеется еще одно фундаментальное ограничение: продолжительность человеческой жизни. Даже самые плодовитые авторы не могут написать так много.


*Издательство* **МИР** *предлагает:*

*К. Бургер*

**СОЛЬВАТАЦИЯ, ИОННЫЕ РЕАКЦИИ И КОМПЛЕКСООБРАЗОВАНИЕ В НЕВОДНЫХ СРЕДАХ.**

Перевод с английского

Книга известного венгерского ученого К. Бургера, неоднократно издававшегося в нашей стране, позволяет читателю ориентироваться в обширной литературе по катализу органических реакций, протекающих под влиянием металлов в неводных средах.

Содержание: Введение. Общая характеристика растворителя. Взаимодействие растворитель — растворенное вещество. Донорно-акцепторное взаимодействие. Донорная и акцепторная способности растворителя. Экспериментальные методы, используемые при изучении неводных растворов сложных систем. Влияние растворителя на структуру и устойчивость металлокомплексов в растворах. Влияние растворителя на кинетику и механизм реакции координации. Взаимодействия в смесях растворителей и их исследование. Общие положения при описании неводных систем. Перспективы области.

Для научных работников — специалистов в области катализа, аналитической химии, неорганической химии.

1984, 22 л. Цена 3 р. 60 к.

*Л. Герфорт, Х. Кох*

**ПРАКТИКУМ ПО РАДИОХИМИИ И РАДИОАКТИВНОСТИ**

Перевод с немецкого

Лабораторное руководство, написанное известными специалистами в области радиоактивности, ведущими также большую преподавательскую работу. В книге не только ставятся и решаются задачи обучения, но и на многочисленных примерах демонстрируются возможности и преимущества использования радиоизотопных методов в различных областях знания.

Книга представляет собой совершенно новое и оригинальное издание, сохранившее от «Практикума по радиохимии» тех же авторов (М.: ИЛ, 1963) его достоинства — логичность построения и изложения, типичность материала. В книге отражены успехи мировой, в том числе и советской науки в развитии научных основ радиоактивности, измерительной техники и радиохимии.

Для преподавателей, аспирантов и студентов химических вузов. Ряд разделов, связанных с новыми методиками и аппаратурой, будет полезен для специалистов, работающих в области атомной энергетики, охраны окружающей среды и других отраслях народного хозяйства.

1984, 36 л. Цена 2 р. 20 к.